Statistical Methods in Medical Research 1995; **4**: 293–309

Design issues in case-control studies

Sholom Wacholder National Cancer Institute, Bethesda, Maryland, USA

The most difficult and most important considerations in planning the protocol of a case-control study are ascertainment of cases, selection of controls and the quality of the exposure measurement. Plans to ensure careful field work are equally important; without attention to data collection, the protocol will be meaningless. In most case-control studies, the measurement problem is magnified because one cannot implement the collection of exposure information at the beginning of follow-up, and instead must rely on interviews, existing records or extrapolation into the past. Consideration of a case-control study as an efficient way to study a cohort helps to resolve some design issues.

A case-control study can be an efficient design for studying the relationship between an exposure and a disease.[1] Instead of selecting individuals for study independently of disease and exposure status, or based on exposure status, the effort in a case-control study is focused on *cases*, those who develop disease, and on a comparison set of *controls*. When the disease is rare, the case-control design achieves a large savings in the number of subjects for whom ascertainment of exposure is required, with only moderate loss of efficiency for estimating a rate-ratio or relative odds compared with a full cohort study, where exposure information is available for everyone.

A case-control study should always be considered in reference to the corresponding full cohort study[1,2] that might have been undertaken in the same *study* base. The study base is the set of individuals in the study population during the time period when they would become cases if they develop disease[3]; thus the unit for measuring the size of a study base is person-time, not persons. The term 'study base' reflects the dynamic nature of the cohort, whose members may enter or exit the study base as their eligibility status changes, whether in a full cohort or a case-control study. The relationship of the case-control study to its underlying cohort through their common study base is central to resolving several tricky design issues discussed below, including control selection and handling of time-dependent covariates, as well as to its analysis and interpretation.

In chronic-disease epidemiology, the parameter of interest is typically the ratio of incidence rates, or events per unit of person-time, at varying levels of exposure. As a consequence, cases will be drawn from disease *incident* during the study period. Anyone with *prevalent* disease should not be included in the study base[4,5] unless a recurrence would be considered a case. Of course, specific objectives of a study may affect these considerations; in a study of birth defects identified in newborns, the relative odds, whose interpretation does not incorporate a time element, would be a more pertinent parameter.

The purpose of this paper is to review the issues that need to be considered when planning a case-control study. The emphasis is on the practical aspects of the design. Sample size and the use of validation studies are discussed in other papers appearing in this issue.

Address for correspondence: Sholom Wacholder, Epidemiology and Biostatistics Program, National Cancer Institute, EPN 403, 6130 Executive Blvd., Rockville, MD 20892-7368, USA.

1 Is a case-control study the design of choice?

The first design issue is deciding whether a case-control study is in fact the best choice. Considerations of data quality, cost, statistical efficiency and credibility will be the dominant determinants of which study design is most appropriate.

1.1 Data quality

In some contexts, the short interval between the launch and the completion of the study is a major advantage of the case-control design. But the validity of a case-control study can be compromised when the short duration is achieved by retrospective collection of exposure information. The possibility of *differential misclassification*, where the reporting of past exposure is affected by the presence or history of disease, is an important concern. While differential misclassification has been difficult to demonstrate, it often cannot be ruled out,[6–10] and is very difficult to correct.

1.2 Cost and statistical efficiency

When will a case-control study be more efficient than a cross-sectional study, where all subjects in the study population are included, and then a cohort study where subjects are chosen for the study on the basis of the likelihood of exposure?

The choice of the most efficient design depends on the fraction of subjects in the study population who develop disease and the fraction who are exposed. In the most interesting situation, both fractions are small. Since the variance of the logarithm of the odds ratio estimate in the standard fourfold table is the sum of the reciprocals of the expected values,[1] the precision of the estimate is most sensitive to the lowest expected number, which will be the expected number of exposed cases. Assume further that exposure ascertainment for an individual is no less expensive than disease ascertainment and would cost the same for any design. An efficient design, therefore, will obtain a fixed number of exposed cases for a relatively small effort. When the proportion of the study population with disease is lower than the proportion exposed, searching for the exposed cases among the cases is more productive than searching for the exposed cases among the exposed. In more complex situations, as when level of exposure is important, the observations that are most influential for estimating the exposure–disease relationship will also be the exposed cases, so, the same design will also be efficient for these purposes.

Cost can modify the conclusions of these simple efficiency considerations. A case-control study would not be appropriate when exposure information, such as from work records, is available easily for everyone contributing to the study base, while a major effort is required to determine disease information. A full cohort study or one that restricted disease follow-up to a subset of the individuals contributing to the study base are superior alternatives. On the other hand, the cost of a cohort study can rise if repeated efforts are required to collect time-dependent variables, in contrast to a case-control study, where all exposure data can be collected at once, albeit retrospectively.

These cost calculations apply most directly to a study designed to evaluate a single agent as a risk factor for a single outcome. When the goal is identification of possibly multiple effects of a single agent, a cohort study may be more appropriate, while a case-control design may be favoured when the purpose is to study several possible risk factors of a single disease.

1.3 Other considerations

Estimates of absolute rates of disease cannot be obtained from case-control studies unless there is a full roster of the underlying cohort[11,12] or the crude incidence rates are available.[13,14] Until recently, only relative measures of association could be estimated; now, the risk difference and other nonmultiplicative parameters can also be estimated.[15] Further, error correction techniques that were developed for cohort studies can be used for case-control studies.[16]

It is also prudent to consider whether the scientific audience for the study is inclined to accept results from any observational study, or particularly from a case-control study.[17] Scepticism of the validity of case-control studies is especially strong in specific areas, such as screening.[18]

2 Choice of setting

Choice of setting should be an important design decision in any study. Considerations include the quality of information on exposure and disease and the costs of collecting them; the variability of the exposure of interest, a factor that directly affects the precision of the estimate; the incidence rate of disease, since a high rate will reduce the length of time needed to accrue a fixed number of cases; likely participation rate[19]; homogeneity with respect to a major confounder that might be difficult to measure or control for, though sometimes an appropriate control group can alleviate the confounding[20,21]; availability of a roster, as in a health-maintenance organization, from which to identify cases and select controls[22,23]; and sociological knowledge of the study setting, e.g. whether the investigators will be able to appraise the likelihood that respondents from an unfamiliar culture are likely to provide honest answers to sensitive questions. Of course, sometimes a particular setting is chosen in order to try to explain extreme rates[24] or geographical or racial differences among rates.[25–27]

2.1 Temporal perspective

All case-control studies have a *retrospective* element to them, since selection based on disease status requires that some actions await determination of an individual's disease status. The temporal relationship between the investigators and the study,[28] i.e. whether the study collected cases as they occur or after the fact is not fundamental, although it is important for case and exposure ascertainment.[29] The prospective approach allows for 'pro-active' determinations of disease and collection of necessary records or materials and reduces the possibility that disease status is affecting a measure of exposure or response to questions; of course, the cost of obtaining exposure information from the full cohort may be prohibitive, even if only storage of materials is required. When thorough medical history or occupational exposure records are available but costly to abstract, a case-control study can be substantially more economical than the corresponding *retrospective full cohort study*.[30]

2.2 Sources of data

2.2.1 Disease data

Incomplete case ascertainment can lead to bias when the exposure of interest is related to the probability of being included.[21] Finding all cases is a major difficulty in many case-control studies. A reliable registry, such as SEER in the USA or the national cancer registries in Scandinavia, can ensure that a high proportion of eligible

cases will be available for study and that misdiagnosis will not corrupt the case series. Quick reporting of disease development to investigators is almost always advantageous. It can reduce the number of proxy respondents needed for rapidly fatal or degenerative diseases and the amount of time when post-disease changes can influence responses to lifestyle questions in interviews. Reliance on other sources, like death certificates, case records, or more informal sources such as a disease-support group[31] may result in less than full ascertainment and some false positives.

2.2.2 Exposure data

Exposure ascertainment can be the trickiest part of a case-control study. Efforts to reduce error by improving the measurement instrument can lead to savings by reducing required sample size.[32] A setting that allows a study to rely on work or medical records rather than on interview data can reduce the likelihood of differential misclassification, though not necessarily of *non*differential misclassification. When self-report is required, a setting where subjects are likely to be co-operative and honest in responding to questions is essential.

2.3 Variability of exposure data

In any study design, including case-control, the variability of the exposure under study is a major determinant of the precision of the estimate of its effect on disease for a given sample size. For exposures that are generally rare, the objective is to find a setting where the exposure is relatively common. For dietary and other continuously-measured exposures, such as consumption of saturated fat in studies of heart disease, the key is the spread, as measured by the variance.

High unconditional variability of exposure is not helpful if variability *conditional on variables that must be stratified for or adjusted on* is substantially lower. Thus, when attempting to disentangle the effects of antihypertensive drugs and hypertension as risk factors for renal cell cancer,[33] the ideal setting would be one where other treatments for hypertension and other indications for drugs used as antihypertensives were common. For a study of the effects of paternal age on the risk of Down syndrome,[34] where maternal age is a known important risk factor, a setting where there tended to be larger discrepancies in both directions between the parents' ages would be better than one where the ages of the two parents tended to be similar.

2.4 Length of follow-up

The precision of estimates of effect from a case-control study increases as the number of cases increases. Just as in a full cohort study, the number of cases can be increased by additional person-time of follow-up from either expanding the study population, or increasing the length of the ascertainment or the accrual period. Possibly, the follow-up period can begin before the study itself, as in a retrospective cohort study. Considerations of sample size and numbers of person-years required to achieve a specified number of cases are discussed in another paper in this issue.

3 Selection of cases and controls

3.1 Case selection

A case series including *all* occurrences of disease in the study base would increase precision of estimates from the study and eliminates the possibility of preferential

selection of cases by severity of disease or by exposure status. When complete ascertainment is impractical or impossible, and particularly for conditions that do not require medical attention such as infertility, the ideal study base would be defined to include only those whose condition would become known if they would develop disease.[21]

Typically, emphasis is placed on documenting that the case developed the disease of interest, apparently following the logic of disease outbreak epidemiology or of the experimental paradigm. Certainly, one would never want to include a case who can be proven not to have the disease of interest. But exclusion of all cases whose diagnosis is not definitive can result in incomplete case ascertainment, reducing precision and leading to possible bias. For example, a protocol that excludes cancer cases for whom there is no histologic confirmation could result in underrepresentation from smaller, rural hospitals and, therefore, bias for an exposure that is related to an agricultural exposure.[35] The most rigorous insistence on confirmation of case status is not always appropriate for a case-control study.

3.2 Principles of control selection

Wacholder *et al.* identified three principles of control selection, which we termed *study base*, *deconfounding* and *equal-accuracy*.[21] They pointed out that efficiency considerations must temper complete compliance with the three comparability principles.[21,36]

3.2.1 The study-base principle

The study base for a case-control study is the same as for the corresponding full cohort study. Ideally, cases would be all or a sample of those in the study base who develop disease under study. Controls can be selected randomly from the study base or chosen so that the distribution of the exposure in the controls is the same as in the study base; in statistical terms, the objective is to be able to consider noncases who are not selected as controls to be *missing* at *random*[27] from the set of eligible noncases.[15]

Miettinen introduced a useful distinction between studies with a *primary base* and those with a *secondary* base.[3,21,38] In a primary-base study, such as the venerable *population-based case-control study*, membership in the study base is relatively easy to ascertain; the challenge is to find the cases that occur within the study base. A secondary-base study begins with available cases and relies on controls to characterize the study base from which those cases arose. The study base is implicitly defined to include anyone who would have become a case *in the study* upon development of the study disease. This defines away the difficulty of cases ascertainment; instead the problems are identifying who is in the study base and choosing a mechanism for selection of controls from whom characteristics of the study base can be extrapolated. An example is a hospital-based study where cases and controls are patients presenting at a same hospital with specified conditions.

The study-base principle is satisfied when cases and controls each constitute a random sample (or a complete enumeration) of incident disease and the nondiseased within the study base. But, while useful conceptually, practical application of the study-base principle can be a challenge.[21,35] Particularly for secondary-base studies, it can be difficult to identify with certainty whether a particular individual is in the study base and eligible to be a control; it is usually impossible to ascertain whether that person would have gone to a study hospital if the disease were diagnosed on a given

day. Use of a particular control series should be recognized as an assumption about the catchment area for the case and control diseases. In practice, of course, sampling of controls from a primary base is also problematic, as will be discussed below.

Occasionally controls may be selected from outside the study base. Female controls could be used in a study of the effect of blood type in males if the blood-type distribution in those females is the same as in males in the study base.[21,39]

In a cohort study, the same criteria are used to decide whether the person-time and incidence of disease from an individual at a given time are included in the study. By analogy, the study-base principle implies that the mechanisms used to identify cases in a case-control study, are intrinsic determinants of proper control selection. It follows that exclusion criteria for cases and controls should be identical.[21,40]

3.2.2 The comparable accuracy principle

When the accuracy of reports by cases is less than perfect, the *comparable accuracy principle* calls for choosing controls so that information from cases and controls is equally reliable, even if controls who would be *more accurate than the cases* are available. For example, selection of controls with another condition has been advocated[6] for a study of a congenital malformation, on the grounds of comparable accuracy. Some reservations about the universal application of this principle are expressed below.

3.2.3 The deconfounding principle

Controls can also be selected to reduce the possibility of confounding.[20,21] When practicable, the most effective way to control for confounding is to restrict the study base to a single level of the confounder. Stratified selection of controls can be particularly useful in a study where it is difficult to control in the analysis alone for an important confounder and where the risk of disease varies by level of that variable for reasons that are difficult to characterize; examples of these confounders include geographic and familial variables. Thus, control of confounding by genotype can be achieved in a study base consisting entirely of strata consisting of two identical twins. It is unclear how effective other kinds of relatives might be as controls, since some variability in the confounding variable will remain even within matched sets. Similarly, the usefulness of matching on neighbourhood or the first eight digits of a ten-digit telephone number, a common practice in random digit dialling (RDD), is also uncertain. In a recent study,[41] this kind of matching in RRD selection produced sets that were close geographically and mostly concordant for socioeconomic variables; it is impossible to determine whether matching on *un*measured variables is achieved. Of course, these kinds of confounding cannot be controlled in the analysis by standard methods either. In all cases, the efficiency and efficacy of using control selection to help eliminate confounding needs to be compared with alternatives that do not involve control selection.

3.3 Roster or no roster

3.3.1 Studies with a roster

A crucial dichotomy from the design perspective is whether or not a *roster* of eligible subjects, i.e. ones who would be cases in the study if they developed disease, is available. Use of a roster of the study base makes control selection much simpler. The roster provides a sampling frame from which to select controls and a list of subjects to check for occurrence of disease; when temporal factors are deemed important, a roster

with dates of beginning and ending of follow-up available makes precise control for the effects of time simpler.

There are several options for selecting controls when a roster is available. The simplest in concept and implementation is the *case-cohort study*,[42] wherein a random sample of *all* eligible subjects in the roster, including cases, are selected as controls. In the *nested case-control study*,[43] whose development antedates the case-cohort study and which remains much more common, separate sets of controls are selected randomly from the members of the roster at risk at the time of each case's development of disease. Each selection should be independent of whether the subject is a future case or a control assigned to another case.[30,44] The sampling can be with replacement if the case is deemed eligible to be selected as its own control, or without replacement otherwise.[45]

Notwithstanding the differences in design, the analyses of case-cohort and nested case-control designs are similar in form[30]; both aim at estimating the hazard or rate ratio comparing the incidence of disease at various levels of exposure. The primary advantage of the nested case-control study seems to be greater efficiency,[30,46] particularly for estimation of time-independent variables. Several practical advantages of the case-cohort design are consequences of its simplicity[47]: use of the same controls for several case groups to overcome the loss of efficiency that would accrue in a study of a single disease; control selection beginning before the roster is identified completely and without risk of needing to discard controls after a putative case is found to be ineligible, perhaps after histologic review; flexibility in choosing among different time-scales, in contrast to the nested case-control study where the time-scale is determined by the sampling scheme; simple external comparisons[48]; and use of the randomly sampled subset of the cohort for learning about characteristics of the cohort.[47]

3.3.2 Primary-base studies without a roster

Ideally, in a case-control study, the case series consists of all cases that arose in the study base, and the controls are selected to be a random sample, perhaps stratified, of the study base. Determination of the sampling frame is problematic without a roster. Schemes such as random digit dialling[49–51] and area sampling[52,53] are expensive attempts to characterize the study base at the time that the cases are diagnosed. Investigators typically are not too satisfied with these, and, therefore, almost universally use alternatives, such as the file from the Health Care Financing Agency for persons aged over 65 in the USA, when available.

These methods might be effective when almost everyone who is eligible has an equal chance of being selected and the response rates are high; the studies are population-based in name only, otherwise. The assumption that probability of exclusion is not related to exposures that are functions of either an individual's lifestyle or that vary geographically is always difficult to justify, so complete coverage and low refusal rates are important. Hartge and Cahill[19] tabulate response rates from studies in the USA from 1979 to 1990; they note evidence of variability, rather than a pronounced decline over time. Perhaps improved fieldwork has compensated for the general decline in response rates to all surveys. Changes in technology and telephone use, such as answering machines[54] and cellular phones, are presenting new challenges for random digit dialling.[35]

Usually, current or past cases are excluded from the controls. In the case-base study,[55,56] controls are sampled from the study base, regardless of their disease status.

This approach allows estimation of the risk-ratio as well as the odds ratio.

3.4 Secondary-base studies

Selection of controls in a secondary-base study can be even trickier. There are two problems: determining who is in the study base (i.e. would be ascertained as cases had they developed disease) and obtaining an appropriate subset from that study base. Often, the simple expedient of choosing patients seen at the study hospital for a different disease is used. If the defining condition for controls is unassociated with the exposure of interest, the exposure distribution in these *hospital controls* will have the same exposure distribution as the study base under the assumption that the catchment populations for the two conditions are the same.[35] Similarly, choosing controls from those with another condition identified in the same way as the case, such as cancer or registry controls, can be reasonable.[57]

As an example, consider an ongoing study of brain cancer at three tertiary-care hospitals (personal communication, P. Inskip *et al.*). Meningiomas and gliomas either diagnosed initially or treated within 30 days of initial diagnosis at each hospital are cases; controls matched to cases by distance between residence and hospital are patients with newly diagnosed conditions seen at neurology, neurosurgery or general surgery wards. These controls are likely to have similar referral patterns to the cases and also offer enough variety to leave reasonable numbers for every important exposure after controls seen for conditions related to that exposure are excluded from analyses.

Other types of controls can be considered. Controls can be selected from the same medical practice as the case on the assumption that all patients originating from the same practice would traverse the medical system in similar ways.[35] Users of medical care controls need to consider the possibility that the practice selectively includes patients with related diseases, potentially leading to a distorted exposure distribution or to changes in the exposure of interest.[58,59] Another problem can be the logistical difficulty of selecting randomly from the practice. A *friend control* group is another example where nonrandom selection can be problematic, particularly since the friends are identified by cases[35]; bias can result in studies of factors related to sociability.[60] On the other hand, the set of potential *relative controls* seems more likely to be completely ascertained. Still, in theory, it is important to divide the study base into mutually exclusive strata and choose friend or relative controls from the case's stratum.[61]

3.5 Which type of control is best?

It is difficult to make general pronouncements about which type of control group is best. A perfect control group would only include those who would be cases if they had developed disease. Thus, a roster simplifies case ascertainment and sampling for controls. But population controls may not be appropriate when only a subset of cases are identified,[31] when recall bias is an issue, or when participation rates are likely to be low.[35] Since it will be almost impossible to state confidently that the three principles delineated above are simultaneously satisfied for any given set of controls,[36] the issues become: which of the practical control series is best, and what is the bias from using the less than perfect control group?

4 Which variables to measure?

Collecting information on variables unnecessarily can be expensive and might detract from the effort on the primary exposure. For example, in a questionnaire study, adding

more questions may reduce the participation rate and the quality of the responses. Further, eventual adjustment for poorly-measured confounders can lead to more bias than is present in the crude estimates.[62] In exploratory studies, where the goal is to identify unknown risk factors, there is clear motivation to collect more variables. Still, though epidemiologists may be reluctant to admit it, there is a point of diminishing returns.

The main objective of the study should determine the variables on which to collect data. When the goal is to evaluate a specific hypothesis, only the exposures of interest, confounders likely to lead to *major* bias and strong effect-modifiers need to be collected. An appropriately conservative rule would be to include *established* risk factors and modifiers of the effect of the exposure of interest and *possible* risk factors likely to be strongly associated with the study exposure. Despite the concern about confounding in the literature,[63–65] it is unusual for failure to control a factor that does not satisfy this rule to affect the interpretation of a study.[66–68]

Collecting greater detail on a confounder can reduce *residual confounding* as noted by Breslow and Day[1] for smoking in a study of lung cancer. But in other contexts with weaker confounders, the extra effort may be unnecessary. Alternatively, a random subset of subjects can be asked about a confounder in detail, and the other subset can be asked in a more simplified form.[69]

While effect-modification is often a major concern, there is no point in collecting expensive data on these variables unless there is sufficient power to detect an important interaction.[70]

5 Data collection

5.1 Maintaining quality

The validity of any epidemiologic study depends fundamentally on fieldwork. Poor fieldwork can undermine the best protocol. Eligible people with disease who are not included as cases; cases who did not actually develop the disease of interest; subjects who are included but not eligible or who are eligible but are improperly excluded, or do not respond or participate can lead to distortion of study results. As long as selection is independent of exposure of interest, conditional on stratification and adjustment variables, there is no bias, however. An effort that succeeds in increasing the response rate could, in theory, cause more harm than good if initial response were independent of exposure and the extra respondents tend to have a different exposure distribution; perhaps more likely, the final set of participants might be a better choice than the early respondents alone. In one study,[71] there was little evidence of difference between early and late respondents to mailed anonymous questionnaires about sexual behaviour.

Often, the most important aspect of an epidemiologic study is the instrument used to assess exposure. This is particularly true in nutritional and environmental studies where self-report is the sole source of exposure information. The ultimate goal should be accurate reports rather than nondifferentially inaccurate ones. Thus, efforts to eliminate report (recall) bias might focus on questionnaire design rather than use of controls with diseases that might have errors similar to those of the cases.[6]

Some case-control studies do not require contact with individuals, but are based solely on medical, occupational or other records. For example, pharmacy records from a health maintenance organization have been a useful resource for studies of side-

effects of prescription drugs.[72] While errors in records are likely to be closer to nondifferential, they must be regarded as fallible.[62] For example, patients do not necessarily use the drugs that their doctors' records indicate were prescribed.[73]

Hartge and Cahill[19] review the state-of-the-art in the conduct of fieldwork. They emphasize the importance of pretesting instruments and procedures; of making the effort to maximize participation; and of monitoring each component of the study on an ongoing basis, particularly if the principal investigator is not directly involved in them.

Fieldwork is one area where case-control studies could use improvement. For example, a cognitive approach to improve the design of data collection instruments has been suggested by Friedenreich.[74] Validation and calibration studies, beyond their use in an attempt to quantify the effect of measurement error, can help to assess the usefulness of the current instrument and to improve the quality of future instruments. Improved approaches to maximize participation rates,[75] increase the quality and completeness of self-reports,[76] prepare questionnaires,[77] and accurately describe the timing of events[78] would be especially helpful.

It is very difficult for a reader to evaluate the quality of data collection. Given its crucial role in epidemiologic studies, much more effort is needed in how to describe and perhaps standardize fieldwork and to assess efforts to improve quality. Quality control techniques can help to identify problem areas; for example, listening to tape-recorded interviews can identify problems not apparent from written materials and in refining questions for subsequent studies.[79]

5.2 Equal accuracy for cases and controls

When the error structure is known, a validation study can sometimes assess the effects of error. Otherwise, there is a strong possibility of bias from the ubiquitous errors in measuring exposures and confounders. When, as is often suspected, the cases tend to overreport or controls tend to underreport exposure, estimates of odds ratios almost always tend to be biased upward. As a result, standard dogma, reflected in the *equal-accuracy principle*, has been that it is advisable to design a study so that controls' and cases' accuracies are similar, in order to avoid a spurious result. There are several reasons to question this view: first, recent work has shown that bias may exaggerate or reverse a relationship rather than attenuate it[62,80–83]; secondly, sometimes errors in controls that match the errors in the cases can increase bias[21]; thirdly, errors in covariates do not generally lead to bias towards the null[62]; fourthly, philosophically, it seems inappropriate to increase error in the name of equality between groups.

5.3 Time-dependent variables and reference dates

Planning for the collection of *time-dependent covariates*, variables that change unpredictably with time, is one area in the design of case-control studies that needs more attention. Two important questions are: In reference to how many different times should the information be collected? and, How should these reference times be determined? A more complete history allows, in principle, assessment of the time during which exposure must occur to convey risk and removes the possibility of not having exposure data at that time. However, the typically high correlation between exposures at two different times means that disentangling latency and similar temporal effects of exposure usually requires a large study with accurate assessment of the changes in exposure. Unfortunately, the extra burden in collecting the additional

detail may reduce its accuracy and increase expense.

One option is to confine the study to those with the least within-person variability. In an ongoing study of the effects of residential electromagnetic fields and childhood leukaemia, the investigators chose to restrict attention to the subjects who moved less often (RL Kleinerman *et al.*, in preparation). This strategy reduces costs, minimizes the problem of integrating field measurements from two or more homes, and obviates the need precisely to specify the time when exposure possibly confers risk; however, it also substantially reduces the opportunity to use the data to determine when that time might be.

How should the reference data be determined, particularly if there is only one period for which exposure assessment is attempted? Just as calendar time and age are options for the primary time scale in the analysis of a cohort study, so too data collection for a time-dependent variable from a matched control can refer either to the date of the diagnosis of the case or to the age at which the control reached the age of diagnosis of the case. The appropriate choice should depend on whether changes in the exposure are more striking with age or with calendar time; for example, use of a personal computer has changed profoundly in the past decade, so a reference date based on calendar time might be preferable, while an age-based reference date might be appropriate for variables such as use of childcare outside the home.

This issue is even more complex in many common circumstances. Choice of reference date for studies becomes more problematic in studies without individual matching. Also, for many diseases, such as prostate, breast, or endometrial cancer, the date of clinical diagnosis depends on nonaetiological factors, possibly resulting in a rather arbitrary reference date. Further, notwithstanding an investigator's valiant efforts, it is difficult for a respondent to report the level of exposure at any point in the past precisely, especially as the elapsed time increases.[74,84] The errors in reporting exposures at two distinct times are likely to be correlated, making disentanglement of temporal effects even more difficult.[62]

5.4 Proxy interviews

In some studies the death or unavailability for interview due to cognitive or communication disorders will be unavoidable for many cases and even some controls.[85] Proxy respondents are less accurate than direct interviews[86,87]; still, use of proxy controls when direct interviews are available in the name of equal accuracy should be discouraged unless the key study variable is prone to considerably more error in the proxy cases than in self-reports from controls.[36]

5.5 Avoiding lengthy questionnaires

Lengthy questionnaires probably reduce participation and accuracy of reports and increase the cost of the study. Thus, trying to obtain more kinds of information and more details about time or level of exposure have subtle negative consequences that must be weighed against the obvious advantages. The *partial questionnaire*[69] is an attempt to preserve the efficiency of a study while reducing the average length of questionnaires by asking about variables of secondary importance to only a randomly-selected fraction of the subjects.

6 Efficiency

Just as considerations of efficiency, a measure of the precision of an estimate of effect attainable for a fixed cost or sample size will often be the overriding determinant of the decision to choose a case-control study, so too will they constrain the options available to investigators in designing the study itself.

6.1 Matching

Matching has been commonly used in studies with and without a roster. Controls are selected so that the value of a covariate, believed to be a confounder, is the same for the case and the controls. The main advantage of matching is the additional efficiency that can sometimes be achieved relative to random sampling when the control and case distributions are substantially different. But the efficiency advantages for matching are often too slight[88] to compensate for: any additional cost or extra effort required to identify controls[89,90]; possible exclusion of cases for whom no match is found[89]; and reduced flexibility in the analysis.[36] Less often, and not always successfully, matching is used in an attempt to capture a set of unmeasured risk factors, such as social class or access to a particular health care facility, in a single variable that is easy to measure, such as neighbourhood.[36]

Sometimes matching hurts rather than helps. *Overmatching* is the term for counter-productive matching, i.e. matching that can cause bias or reduce precision.[36,91,92] Matching on a variable in the pathway between exposure and disease can lead to bias. An example would be matching on endometrial hyperplasia in a study of oestrogen and endometrial cancer.[91,92] Matching on a variable that is not itself a strong risk factor can lead to reduction in precision if it reduces the variability of the exposure *conditional on the matching variable*, i.e. the variability that is a strong determinant of the precision of the estimate of effect. Finally, the analysis of a matched study needs to account for the matching, in contrast to unmatched studies, where a decision about stratification can be made at the analysis stage.

6.2 Two-stage designs instead of frequency matching

Two-stage designs include sampling schemes where the value of an easy-to-obtain first-stage variable X_1 determines the probability of selection for determination of the expensive second-stage variable X_2. For example, Weinberg and Sandler[93] discuss a two-stage design where the objective is to examine the joint effects of smoking (X_1) and indoor radon exposure (X_2). The appropriate two-stage design will depending on specifics of the design and on whether interest lies in the main effect of X_1 or X_2 or in their interaction.[93–96].

Frequency matching can be considered to be a special case of a two-stage design. A major advantage of some of the two-stage approaches is that all the available data on X_1 is used, in contrast to matching where the values of matching variables for potential subjects who fail the matching test are excluded from the analysis. Two-stage designs have several other advantages over matched studies: ability fully to manipulate the probabilities of selection of both cases and controls for measurement of X_2 on the basis of X_1, rather than only selecting controls to match the observed distribution of X_1 in cases; ability to estimate main effects of X_1 as well as X_2; additional power for main effects of X_2 and for $X_1 - X_2$ interactions[94]; and more flexibility in the analysis[97] of the joint effects of X_1 and X_2. One of the two-stage designs will usually be a better option than frequency matching because of the more flexible design and analysis. Aside from

the lack of easily usable software for analysis, the one disadvantage of most two-stage designs is the requirement that eligibility criteria be verified, even for those for whom X_2 is not measured.[97] One two-stage design, *randomized recruitment*,[95] however, does not require eligibility verification, and may be appropriate, for example, when a detailed residential history is required.

6.3 Multiple case and control groups

A single control group can be used for more than one case group. In addition to the obvious advantage of efficiency, the estimates of effect for the various disease groups will be correlated, providing some calibration of the control group.[36] Multiple control groups, as has been advocated,[98] does not help with the question of which group is correct when the results are discrepant.[36]

6.4 Case-only designs

When two exposures X_1 and X_2 can be assumed to be independent in the study base, a design with no controls can be used to estimate the ratio of the odds ratios for X_1 at different levels of X_2 (Umbach DM, Weinberg CR, unpublished observations, 1995).[99] The assumption that a genotype and an environmental exposure are independent justifies a case-only analysis that allows estimation of the effect of the exposure and provides a more precise estimate of the interaction on a multiplicative scale than the standard analysis (Umbach DM, Weinberg CR, unpublished observations, 1995). The danger of this design is that the independence assumption cannot be verified internally without studying at least some controls (Umbach DM, Weinberg CR, unpublished observations, 1995).

Variation in time-dependent covariates is linked to time of the event in the *case-crossover* design.[100,101] Behaviour, such as level of physical exertion, is compared in the period immediately before a myocardial infarction and a specified earlier period, such as 24 hours earlier.[102] Here, each individual forms a unique stratum, and those who are 'concordant' with respect to disease, i.e. for whom there is no event during follow-up, are excluded.[35,101]

7 Outlook

The realization that a case-control study and its corresponding cohort study share a common study base provides insight into nearly all the issues in the design of case-control studies. Collection of exposure information, particularly when it changes with time, is usually the most difficult challenge in the design of case-control studies. Innovations in design and improvements in fieldwork may lead to better, more convincing and more efficient case-control studies.

References

1 Breslow NE, Day NE. *Statistical methods in cancer research, Volume 1, The analysis of case-control studies*, IARC Scientific Publications No. 32. Lyon: International Agency for Research on Cancer, 1980.

2 Breslow NE, Day NE. *Statistical methods in cancer research, Volume 2, The design and analysis of cohort studies*, IARC Scientific Publications No. 82. Oxford: Oxford University Press, 1987.

3 Miettinen OS. *Theoretical epidemiology: principles of occurrence research in medicine.* New York: John Wiley and Sons, 1985.

4 Miettinen OS. Estimability and estimation in case-referent studies. *American Journal of Epidemiology* 1976; **103**: 226–35.

5 Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology* 1982; **116**: 547–53.

6 Werler MM, Pober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. *American Journal of Epidemiology* 1989; **129**: 415–21.

7 Harlow SD, Linet MS. Agreement between questionnaire data and medical records: the evidence of accuracy of recall. *American Journal of Epidemiology* 1989; **129**: 222–48.

8 Mackenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *American Journal of Epidemiology* 1989; **129**: 65–75.

9 Friedenreich CM, Howe GR, Miller AB. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. *Annals of Epidemiology* 1991; **1**: 439–53.

10 Giovanucci E, Stampfer MJ, Colditz GA, *et al.* A comparison of prospective and retrospective assessments of diet in the study of breast cancer. *American Journal of Epidemiology* 1993; **137**: 502–11.

11 Thomas DC, Blettner M, Day NE. Use of external rates in nested case-control studies with application to the international radiation study of cervical cancer patients. *Biometrics* 1992; **48**: 781–94.

12 Borgan Ø, Langholz B. Non-parametric estimation of relative mortality from nested case-control studies. *Biometrics* 1993; **49**: 593–602.

13 Greenland S. Multivariate estimation of exposure-specific incidence from case-control studies. *Journal of Chronic Diseases* 1981; **34**: 445–53.

14 Benichou J, Wacholder S. A comparison of three approaches to estimate exposure-specific incidence rates from population-based case-control data. *Statistics in Medicine* 1994; **13**: 651–61.

15 Wacholder S. The case-control study as a missing data problem: estimating the risk difference. *Epidemiology* 1995 (in press).

16 Carroll RJ, Wang S, Wang CY. Prospective analysis of case-control studies. *Journal of the American Statistical Association* 1995; **90**: 157–69.

17 Feinstein AR. Scientific standards in epidemiologic studies of the menace of daily life. *Science* 1988; **242**: 1257–63.

18 Connor RJ, Prorok PC, Weed DL. The case-control design and the assessment of the efficacy of cancer screening. *Journal of Clinical Epidemiology* 1991; **44**: 1215–21.

19 Hartge P, Cahill JI. Field methods in epidemiology. In: Rothman K, Greenland S eds. *Modern epidemiology.* Boston: Little Brown, 1996 (in press).

20 Wickramaratne PJ, Holford TR. Confounding in epidemiologic studies: the adequacy of the control group as a measure of confounding. *Biometrics* 1987; **43**: 751–65.

21 Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies: I. Principles. *American Journal of Epidemiology* 1992; **135**: 1019–28.

22 Schiffman MH, Bauer HM, Hoover RN *et al.* Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *Journal of the National Cancer Institute* 1993; **85**: 958–64.

23 Boice JD, Morin MM, Glass AG *et al.* Diagnostic X-ray procedures and risk of leukemia, lymphoma, and multiple myeloma. *Journal of the American Medical Association* 1991; **265**: 1290–94.

24 Li JY, Ershow AG, Chen ZJ *et al.* A case-control study of cancer of the esophagus and gastric cardia in Linxian. *International Journal of Cancer* 1989; **43**: 755–61.

25 Hartge P, Cahill J, West D *et al.* Design and methods in a multi-center case-control interview study. *American Journal of Public Health* 1984; 74: 52–6.

26 Silverman DT, Levin LI, Hoover RN, Hartge P. Occupational risk factors of

bladder cancer in the United States. I. White men. *American Journal of Epidemiology* 1986; **123**: 174–84.
27 Brown LM, Hoover RN, Greenberg RS *et al.* Are racial differences in squamous cell esophageal cancer explained by alcohol and tobacco use? *Journal of the National Cancer Institute* 1994; **86**: 1340–45.
28 Kramer MS, Boivin JF. Directionality, timing, and sample selection in epidemiologic research design. *Journal of Clinical Epidemiology* 1989; **42**: 827–8.
29 Greenland S, Morgenstern H. What is directionality? *Journal of Clinical Epidemiology* 1989; **42**: 821–4.
30 Wacholder S, Gail MH, Pee D. Selecting an efficient design for assessing exposure-disease relationships in an assembled cohort. *Biometrics* 1991; **47**: 63–76.
31 Perneger TV, Whelton PK, Klag MG. Risk of kidney failure associated with the use of acetaminophen, aspirin, and nonsteroidal antiinflammatory drugs. *New England Journal of Medicine* 1994; **331**: 1675–9.
32 Walker AM, Blettner M. Comparing imperfect measures of exposure. *American Journal of Epidemiology* 1988; **121**: 783–90.
33 Chow WH, McLaughlin JK, Mandel JS, Wacholder S, Niwa S, Fraumeni JF. Risk of renal cell cancer in relation to diuretics, antihypertensive drugs, and hypertension. *Cancer Epidemiology Biomarkers and Prevention* 1995; **4**: 327–31.
34 Regal RR, Cross PK, Lamson SH, Hook EB. Search for evidence for a paternal age effect independent of maternal age effect in birth certificate reports of Down's syndrome in New York state. *American Journal of Epidemiology* 1980; **112**: 650–55.
35 Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies: II. Types of controls. *American Journal of Epidemiology* 1992; **135**: 1029–41.
36 Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies: III. Design options. *American Journal of Epidemiology* 1992; **135**: 1042–50.
37 Little RJ, Rubin DB. *Statistical analysis with missing data.* New York: Wiley, 1987.
38 Miettinen OS. The 'case-control' study: valid selection of subjects. *Journal of Chronic Diseases* 1985; **38**: 543–8.
39 Miettinen OS, Cook EF. Confounding: essence and detection. *American Journal of Epidemiology* 1981; **114**: 593–603.
40 Lubin JH, Hartge P. Excluding controls: misapplications in case-control studies. *American Journal of Epidemiology* 1984; **120**: 791–3.
41 Sakkinen PA, Severson RK, Ross JA, Robison LL. Random-digit dialing for control selection in childhood cancer studies: the geographic proximity and demographics within matched sets. *American Journal of Public Health* 1995; **85**: 555–7.
42 Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Bi*
43 Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: appraisal by application to asbestos mining (with discussion). *Journal of the Royal Statistical Society A* 1977; **140**: 469–91.
44 Lubin JH, Gail MH. Biased selection of controls for case-control analysis of cohort studies. *Biometrics* 1984; **40**: 63–75.
45 Robins J, Gail MH, Lubin JH. More on biased selection of controls. *Biometrics* 1986; **42**: 2[illegible]–9.
46 Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *American Journal of Epidemiology* 1990; **131**: 169–76.
47 Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control design. *Epidemiology* 1991; **2**: 155–8.
48 Wacholder S, Boivin J-F. External comparisons with the case-cohort design. *American Journal of Epidemiology* 1987; **126**: 1198–209.
49 Waksberg J. Sampling methods for random digit dialing. *Journal of the American Statistical Association* 1978; **73**: 40–46.
50 Greenberg ER. Random digit dialing for control selection: a review and a caution on its use in studies of childhood cancer. *American Journal of Epidemiology* 1990; **131**: 1–5.
51 Potthoff RF. Telephone sampling in epidemiologic research: to reap the benefits, avoid the pitfalls. *American Journal of Epidemiology* 1994; **139**: 967–78.
52 Kish L. *Survey sampling.* New York: John Wiley and Sons, 1965.
53 Lele C, Holly EA, Roseman DS, Thomas DB. Comparison of control subjects recruited by random digit dialing and area survey. *American Journal of Epidemiology* 1994; **140**: 643–8.

54 Harlow BL, Crea EC, East MA *et al.* Telephone answering machines: the influence of leaving messages on telephone interviewing response rates. *Epidemiology* 1993; **4**: 380–83.

55 Kupper LL, McMichael AJ, Spirtas R. A hybrid epidemiologic study design useful in estimating relative risk. *Journal of the American Statistical Association* 1975; **70**: 524–8.

56 Miettinen OS. Design options in epidemiologic research: an update. *Scandinavian Journal of Work and Environmental Health* 1982; **8**(suppl 1: 14.

57 Pearce N, Checkoway H. Case-control studies using other diseases as controls: problems of excluding exposure-related diseases. *American Journal of Epidemiology* 1988; **127**: 851–6.

58 Silverman DT, Hoover RN, Swanson GM, Hartge P. The prevalence of coffee drinking among hospitalized and population-based control groups. *Journal of the American Medical Association* 1983; **249**: 1877–80.

59 MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas [letter]. *New England Journal of Medicine* 1981; **304**: 1605–606.

60 Siemiatycki J. Friendly control bias. *Journal of Clinical Epidemiology* 1989; **42**: 687–8.

61 Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. *Epidemiology* 1990; **1**: 273–84.

62 Wacholder S. When measurement error correlates with truth: surprising effects of differential misclassification. *Epidemiology* 1995; **6**: 157–61.

63 Boivin JF, Wacholder S. Conditions for confounding of the risk ratio and of the odds ratio. *American Journal of Epidemiology* 1985; **121**: 152–8.

64 Grayson DA. Confounding confounding. *American Journal of Epidemiology* 1987; **126**: 546–53.

65 Robins J, Greenland S. Identifiability, exchangeability and epidemiologic confounding. *International Journal of Epidemiology* 1986; **15**: 413–19.

66 Yanagawa T. Case-control studies: assessing the effect of a confounding factor. *Biometrika* 1984; **71**: 191–4.

67 Siemiatycki J, Wacholder S, Dewar R *et al.* Degree of confounding bias related to smoking, ethnic group, and socioeconomic group in estimates of the associations between occupation and cancer. *Journal of Occupational Medicine* 1988; **30**: 617–25.

68 Blair A, Hoar S, Walrath J. Comparison of crude and smoking-adjusted standardised mortality ratios. *Journal of Occupational Medicine* 1985; **27**: 881–4.

69 Wacholder S, Carroll RJ, Pee D, Gail M. The partial questionnaire design for case-control studies. *Statistics in Medicine* 1994; **13**: 623–34.

70 Greenland S. Tests for interaction in epidemiologic studies: a review and study of power. *Statistics in Medicine* 1982; **2**: 243–51.

71 Biggar RJ, Melbye M. Responses to anonymous questionnaires concerning sexual behavior: a method to examine potential biases. *American Journal of Public Health* 1992; **82**: 1506–12.

72 Finkle WD, McLaughlin JK, Ragson SA *et al.* Increased risk of renal cell cancer among women using diuretics. *Cancer Causes Control* 1993; **4**: 555–8.

73 Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *American Journal of Epidemiology* 1993; **137**: 1251–8.

74 Friedenreich CM. Improving long-term recall in epidemiologic studies. *Epidemiology* 1994; **5**: 1–4.

75 Groves RM, Lyberg LE. An overview of nonresponse issues in telephone surveys. In: Groves RM, Biemer PB, Lyberg LE, Massey JT, Nicholls WC, Waksberg J eds. *Telephone survey methodology*. New York: John Wiley and Sons, 1988: 191–211.

76 Cannell C, Oksenberg L. Observation of behavior in telephone interviews. In: Groves RM, Biemer PB, Lyberg LE, Massey JT, Nicholls WC, Waksberg J eds. *Telephone survey methodology*. New York: John Wiley and Sons, 1988: 475–95.

77 Sudman S, Bradburn NM. *Asking questions: a practical guide to questionnaire design*. San Francisco: Jossey-Bass, 1982.

78 Auriat N. A comparison of event dating accuracy between the wife, the husband, the couple, and the Belgium population register. *Public Opinion Quarterly* 1993; **57**: 165–90.

79 Edwards S, Slattery ML, Mori M *et al.* Objective system for interviewer performance evaluation for use in epidemiologic studies. *American Journal of Epidemiology* 1994; **140**: 1029–37.

80 Dosemeci M, Wacholder S, Lubin JH. Does non-differential misclassification always bias an association towards the null? *American*

Journal of Epidemiology 1990; **132**: 746–8.
81 Wacholder S, Dosemeci M, Lubin J. Blind exposure ascertainment does not guarantee nondifferential misclassification. *American Journal of Epidemiology* 1991; **134**: 433–7.
82 Flegal KM, Keyl PM, Javier Nieto F. Differential misclassification arising from nondifferential errors in exposure measurement. *American Journal of Epidemiology* 1991; **134**: 1233–44.
83 Weinberg CR. Umbach DM, Greenland S. When will nondifferential misclassification of an exposure preserve the direction of a trend? *American Journal of Epidemiology* 1994; **10**: 565–71.
84 Bradburn MM, Riups LJ, Shivell SK. Answering autobiographical questions: the input of memory and inference on surveys. *Science* 1987; **2236**: 157–61.
85 Nelson LM, Longstreth WT, Koepsell TD, van Belle G. Proxy respondents in epidemiologic research. *Epidemiologic Review* 1990; **12**: 71–86.
86 McLaughlin JK, Dietz MS, Mehl ES, Blot WJ. Reliability of surrogate information in cigarette smoking by type of informant. *American Journal of Epidemiology* 1987; **126**: 144–6.
87 McLaughlin JK, Mandel JS, Mehl ES, Blot WJ. Reliability of next-of-kin and self-respondents for cigarette, coffee, and alcohol consumption. *Epidemiology* 1990; **1**: 408–12.
88 Thomas DC, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. *Journal of Chronic Diseases* 1983; **36**: 685–97.
89 McKinlay SM. Pair matching—reappraisal of a popular technique. *Biometrics* 1977; **33**: 725–35.
90 Walter SD. The feasibility of matching and quota sampling in epidemiologic studies. 1989; **130**: 379–89.
91 Cole P. *Statistical methods in cancer research, Volume 1, The analysis of case-control studies (Introduction) IARC Scientific Publications No. 32.* Lyon: International Agency for Research on Cancer, 1980: 14–40.
92 Breslow N. Design and analysis of case-control studies. *Annual Review of Public Health* 1982; **3**: 29–54.
93 Weinberg CR, Sandler D. Randomized recruitment in case-control studies. *American Journal of Epidemiology* 1991; **134**: 421–32.
94 Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika* 1988; **75**: 11–20.
95 Weinberg CR, W[illegible] and analysis of case-control studies with biased sampling. *Biometrics* 1990; **46**: 963–75.
96 Langholz B, Clayton D. Sampling strategies in nested case-control studies. *Environmental Health Perspectives* 1994; **102**(suppl 8):47–51.
97 Wacholder S, Weinberg CR. Flexible maximum likelihood methods for assessing joint effects in case-control studies with complex sampling. *Biometrics* 1994; **50**: 350–57.
98 Ibrahim MS, Spitzer WO. The case-control study: the problem and the prospect. *Journal of Chronic Diseases* 1979; **32**: 139–44.
99 Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* 1994; **13**: 158–62.
100 Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* 1991; **133**: 144–53.
101 Mittleman MA, Maclure M, Robins JM. Control sampling strategies for case-crossover studies: an assessment of relative efficiency. *American Journal of Epidemiology* 1995; **142**: 91–8.
102 Mittleman MA, Maclure M, Tofler GH *et al.* Triggering of acute myocardial infarction by heavy physical exertion. *New England Journal of Medicine* 1993; **329**: 1677–83.